

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): MURAKAMI, Toshihiko  
Serial No.: Not yet assigned  
Filed: September 17, 2003  
Title: DATA TRANSFER METHOD  
Group: Not yet assigned

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

September 17, 2003

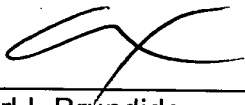
Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on Japanese Patent Application No.(s) 2003-038097, filed February 17, 2003.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP



---

Carl I. Brundidge  
Registration No. 29,621

CIB/alb  
Attachment  
(703) 312-6600

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日            2 0 0 3 年   2 月 1 7 日  
Date of Application:

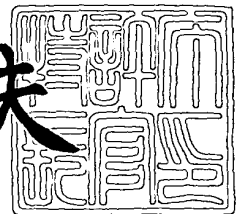
出 願 番 号            特 願 2 0 0 3 - 0 3 8 0 9 7  
Application Number:  
[ST. 10/C]:            [ J P 2 0 0 3 - 0 3 8 0 9 7 ]

出   願   人            株 式 会 社 日 立 製 作 所  
Applicant(s):

2 0 0 3 年   7 月 3 1 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



出証番号   出証特 2 0 0 3 - 3 0 6 1 1 5 2

【書類名】 特許願

【整理番号】 K201032I

【提出日】 平成15年 2月17日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 12/02

【発明者】

【住所又は居所】 神奈川県川崎市麻生区王禅寺 1099番地 株式会社  
日立製作所 システム開発研究所内

【氏名】 村上 俊彦

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100099298

【弁理士】

【氏名又は名称】 伊藤 修

【連絡先】 03-3251-3824

【選任した代理人】

【識別番号】 100099302

【弁理士】

【氏名又は名称】 笹岡 茂

【手数料の表示】

【予納台帳番号】 018647

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 データ移行方法

【特許請求の範囲】

【請求項 1】 複数の計算機と複数の記憶装置と、前記計算機と記憶装置を相互に接続する中継装置と、前記計算機と記憶装置と中継装置を管理する管理装置とで構成された計算機システムにおけるデータ移行方法であって、

前記管理装置は、前記複数の計算機に対して前記記憶装置の仮想的な記憶領域を設定し、該設定の内容の情報を第 1 の情報として保持し、

前記中継装置は、前記第 1 の情報を元に作成された第 2 の情報を保持し、

前記仮想的な記憶領域は前記それぞれの記憶装置内の記憶領域あるいは複数の記憶装置内の記憶領域が結合されてなる記憶領域に対応し、

前記中継装置は、前記第 2 の情報から一つの仮想的な記憶領域を選択し、該選択した仮想的な記憶領域が、複数の記憶装置内の記憶領域が結合されてなる記憶領域である場合を契機として、該複数の記憶装置間でのデータの移行を行い、該結合にかかわる記憶装置を少なくすることを特徴とするデータ移行方法。

【請求項 2】 請求項 1 記載のデータ移行方法において、

前記中継装置は、前記第 2 の情報を参照し、前記仮想的な記憶領域に対応していない比較的小さな容量の記憶領域が増加してきた場合を契機として、該複数の記憶装置間でのデータの移行を行い、前記仮想的な記憶領域に対応していない比較的小さな容量の記憶領域を少なくすることを特徴とするデータ移行方法。

【請求項 3】 請求項 1 または請求項 2 記載のデータ移行方法において、

前記第 2 の情報は前記第 1 の情報を元に更新されることを特徴とするデータ移行方法。

【請求項 4】 請求項 1 または請求項 2 記載のデータ移行方法において、

前記複数の記憶装置間でデータ移行をする際に、データ移行の単位ごとにデータ移行未完かデータ移行済かを示す第 3 の情報を設定し、該データ移行の単位がデータ移行されるごとに前記データ移行未完をデータ移行済に変更し、データ移行の進行状況を更新することを特徴とするデータ移行方法。

【請求項 5】 請求項 3 記載のデータ移行方法において、

前記第二の情報は前記仮想的な記憶領域毎にデータ移行中であるかどうかの状態を示すフラグを有し、

前記中継装置は、前記計算機が前記仮想的な記憶領域をアクセスするときに、前記第二の情報の該当する前記仮想的な記憶領域の前記フラグがデータ移行中である場合は、前記第三の情報によりデータ移行元またはデータ移行先のどちらにアクセスを行うべきかを判断し、データ移行中に前記計算機から前記仮想的な記憶領域へのデータのアクセスを中断させないことを特徴とするデータ移行方法。

【請求項 6】 請求項 1 乃至請求項 5 のいずれかの請求項記載のデータ移行方法において、

前記中継装置は、前記仮想的な記憶領域の一部であるデータ移行を行う記憶領域のデータを、データ移行先の記憶領域へ直接コピーすることを特徴とするデータ移行方法。

【請求項 7】 請求項 1 乃至請求項 5 のいずれかの請求項記載のデータ移行方法において、

前記中継装置は、前記仮想的な記憶領域の一部であるデータ移行を行う記憶領域のデータを、データ移行のために予め用意している前記記憶装置内の記憶領域に一旦コピーした後に、データ移行先の記憶領域へ間接的にコピーすることを特徴とするデータ移行方法。

【請求項 8】 請求項 1 乃至請求項 5 のいずれかの請求項記載のデータ移行方法において、

前記中継装置は、前記仮想的な記憶領域の一部であるデータ移行を行う記憶領域のデータを、前記複数の記憶装置の記憶領域の未使用の記憶領域に一旦コピーした後に、データ移行先の記憶領域へ間接的にコピーすることを特徴とするデータ移行方法。

【請求項 9】 請求項 1 乃至請求項 5 のいずれかの請求項記載のデータ移行方法において、

前記中継装置は、データ移行時の前記仮想的な記憶領域の一部であるデータをデータ移行後も、データ移行前の仮想的な記憶領域に一時的にまたは指定した期間残しておくことにより、データ移行後の新しい仮想的な記憶領域のバックアップ

データとすることを特徴とするデータ移行方法。

【請求項 10】 請求項 1 乃至請求項 5 のいずれかの請求項記載のデータ移行方法において、

前記中継装置は、前記仮想的な記憶領域の一部であるデータ移行を行う記憶領域のデータを、データ移行のために予め用意している前記中継装置内の記憶領域に一旦コピーした後に、データ移行先の記憶領域へ間接的にコピーすることを特徴とするデータ移行方法。

【請求項 11】 請求項 1 記載のデータ移行方法において、

前記計算機システム内の前記中継装置を冗長構成とした場合、前記管理装置は、前記第一の情報を全ての前記中継装置へと配布し、前記第二の情報の情報源とすることを特徴とするデータ移行方法。

【請求項 12】 請求項 1 記載のデータ移行方法において、

前記計算機システム内の前記中継装置の内部の構成要素を冗長構成とし、前記第二の情報を持つ構成要素が冗長構成となる場合、当該冗長構成の構成要素間で第二の情報の同期を常にとることにより、当該冗長構成の構成要素の一つが障害となった場合に、その他の当該冗長構成の構成要素の前記第二の情報を前記中継装置が利用することを特徴とするデータ移行方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、複数の記憶装置の記憶領域を仮想化して利用する仮想化技術に関わり、記憶装置間または記憶装置内のデータ移行方法に関する。

【0002】

【従来の技術】

複数の計算機と複数の記憶装置をネットワークで接続する SAN (Storage Area Network) では、記憶装置の記憶領域を仮想化して利用させる技術がある (例えば、非特許文献 1 参照)。

計算機と記憶装置の間のどの位置で仮想化を行うかによりいくつかの方法がある。

第一はサーバアプリケーションが実行される計算機で、ストレージ管理ソフトウェアやボリューム管理ソフトウェアにより仮想化を行う方法である。

第二は記憶装置の手前に記憶装置を接続するインタフェースを複数用意した計算機を置き、この計算機で仮想化を行う方法や記憶装置そのものが仮想化を行う方法である。

第三はSANを構成するネットワーク機器や管理サーバ装置で仮想化を行う方法で、制御用とデータ用のI/O (Input/Output) を同じネットワークで扱うインバンド方法と、制御用のI/Oを流すネットワークを別途設けてデータ用のI/Oとは別に扱うアウトオブバンド方法の二つの方法がある。

【非特許文献1】

Evaluator Group社の技術白書である「Virtualization of Disk Storage (2000年9月)」

【0003】

【発明が解決しようとする課題】

従来の技術で示した第二または第三の仮想化の方法で、ファイバチャネルスイッチやイーサネット（登録商標）でiSCSI (Internet Small Computer System Interface) を使用してSANを構成する場合のLANスイッチのような中継装置や、ファイバチャネルやイーサネット（登録商標）のHBA (Host Bus Adaptor) を複数備えるサーバベースの装置が仮想化された記憶領域（仮想ボリューム）を提供している場合に、仮想ボリュームを複数の記憶装置の記憶領域の結合で構成していると、仮想ボリュームのアクセスの状況によっては、複数の記憶装置の中の、例えば、第一の記憶装置、第二の記憶装置、第三の記憶装置という具合に、次々にアクセスする宛先の記憶装置が変化し、中継処理負荷が増加するという問題が発生する。

本発明の第一の目的は、比較的少ない記憶装置からの記憶領域で仮想ボリュームを構成するように、運用中にデータ移行を行うことにより、中継処理負荷を低減させる方法を提供することにある。

また、複数の記憶装置の記憶領域から仮想ボリュームを構成し、領域の拡大または縮小を随時行っていくと、比較的小さな容量の未使用の記憶領域が増加する

場合が考えられ、多数の小さな容量の記憶領域で仮想ボリュームを構成すると、中継装置における仮想ボリュームと実際の記憶領域との対応の変換処理の処理負荷が増加する。

そして、比較的大きな仮想ボリュームを比較的小さい数の未使用の記憶領域から構成することができなくなるという問題も発生する。

本発明の第二の目的は、比較的小さな容量の未使用の記憶領域が少なくなるように、既に仮想ボリュームを構成している一つ以上の記憶領域を別の記憶領域へと、運用中にデータ移行を行うことにより、この問題を低減させる方法を提供することにある。

#### 【0004】

##### 【課題を解決するための手段】

上記目的を達成するため、本発明は、複数の計算機と複数の記憶装置と、前記計算機と記憶装置を相互に接続する中継装置と、前記計算機と記憶装置と中継装置を管理する管理装置とで構成された計算機システムにおけるデータ移行方法であり、

管理装置は、複数の計算機に対して記憶装置の仮想的な記憶領域を設定し、該設定の内容の情報を第1の情報として保持し、中継装置は、第1の情報を元に作成された第2の情報を保持し、

仮想的な記憶領域は前記それぞれの記憶装置内の記憶領域あるいは複数の記憶装置内の記憶領域が結合されてなる記憶領域に対応し、

中継装置は、第2の情報から一つの仮想的な記憶領域を選択し、該選択した仮想的な記憶領域が、複数の記憶装置内の記憶領域が結合されてなる記憶領域である場合を契機として、該複数の記憶装置間でのデータの移行を行い、該結合にかかわる記憶装置を少なくする。

また、中継装置は、第2の情報を参照し、仮想的な記憶領域に対応していない比較的小さな容量の記憶領域が増加してきた場合を契機として、複数の記憶装置間でのデータの移行を行い、前記仮想的な記憶領域に対応していない比較的小さな容量の記憶領域を少なくする。

#### 【0005】



**【発明の実施の形態】**

本発明の第一の実施形態を図 1 から図 13 を参照して説明する。

第一の実施形態では、比較的少ない記憶装置からの記憶領域により仮想ボリュームの構成を行うように、運用中にデータ移行を行う場合について説明する。

図 1 は、本発明における計算機システムの構成例を示す図である。

サーバ装置 170 がスイッチ 100 を介してストレージ装置 180 と接続される。

スイッチ 100 は、ファイバチャネルスイッチや LAN スwitch であるが、本発明の実施例の説明の中継装置としては、イーサネット（登録商標）Switch の場合を例として示す。

スイッチ 100 は、サーバ装置 170 またはストレージ装置 180 等を接続するための物理層とデータリンク層の制御を行う複数の接続インタフェース 110（図 1 では、Ether MAC と表記している）と、送受信するパケットのレイヤ 2 から上位（図中はレイヤ 7 までと示している）のヘッダまたは情報部分の情報を元に宛先となる接続インタフェースの決定や必要時にはパケットの内容の変換を行う複数のルーティング制御部（Routing Control）120 と、複数のルーティング制御部を接続するクロスバースwitch（Crossbar Switch）130 と、スイッチ 100 の装置管理や経路制御プロトコルの計算を行うスイッチ管理部（Switch Manager）140 と、クロスバースwitch 130 とスイッチ管理部 140 を接続する内部バス等の内部通信線 101C で構成される。

**【0006】**

ルーティング制御部 120 は、CPU 126、プログラムやテーブルを格納するメインメモリ（MEM）127、レイヤ 2 からレイヤ 7 のパケットのフォワーディング制御部（Layer2-7 Forwarding Engine）121、パケットのヘッダ検索部（Search Engine）122 とそのテーブル 125 を格納するメモリ（MEM）123 を有し、これらが内部バス等の内部通信線 128 で接続されている。

スイッチ管理部 140 は、CPU 141、プログラムや経路制御テーブル等のテーブル 143 を格納するメインメモリ（MEM）142、管理用のイーサネット（登録商標）（Ether）145、クロスバースwitch への内部通信線 101C と

接続するための内部バス制御 144 を有し、これらが内部バス等の内部通信線 146 で接続されている。

スイッチ 100 には、さらに管理用のコンソールまたはイーサネット（登録商標）のポート 161 を介して管理装置 160 や、接続インタフェース（Ether MAC） 110 を介してストレージ装置 180 の実ボリューム、またはスイッチ 100 が提供する仮想ボリュームのバックアップ等の機能をスイッチと連携して行うための複数の連携サーバ 150 が接続される。

連携サーバ 150 は、CPU 151、プログラムやボリューム管理テーブル等のテーブル 153 を格納するメインメモリ（MEM） 152、イーサネット（登録商標）（Ether） 156、ハードディスク（hard disk） 157 を有し、これらが内部バス制御 154 を介して内部バス等の内部通信線 155 で接続されている。

#### 【0007】

図 2 は、スイッチ 100 のスイッチ管理部 140 内のメインメモリ 142 にあるプログラムおよびテーブルの構成例を示す図である。

メインメモリ 142 のプログラム領域 201 は、オペレーティングシステム 203、経路制御プロトコル 204、実／仮想ボリューム管理 207、装置管理 209、スイッチ宛パケット処理 211、データ移行処理 212 等から構成され、テーブル領域 143 は、経路制御プロトコル 204 と関係するルーティングテーブル 205 とフィルタリングテーブル 206、実／仮想ボリューム管理 207 と関係するボリューム管理テーブル 208、装置管理 209 と関係する装置管理テーブル 210、ユニット使用状況管理テーブル 213 等から構成される。装置管理テーブルは接続されている装置および内蔵する装置の種類、構成情報、性能情報等を記録したテーブルであり、図 4、5 における装置管理テーブルも同様のものである。

#### 【0008】

図 3 は、スイッチ 100 のルーティング制御部 120 内のメインメモリ 127 およびメモリ 123（125）にあるプログラムおよびテーブルの構成例を示す図である。

メインメモリ 127 はオペレーティングシステム 301、ハード処理不可パケ

ット処理 303、スイッチ管理部との通信処理 302 等から構成され、メモリ 125 は、スイッチ管理部との通信処理 302 と関係するルーティングテーブル 304、フィルタリングテーブル 305、ボリューム管理テーブル 306 等から構成される。

#### 【0009】

図 4 は、連携サーバ 150 のメインメモリ 152 にあるプログラムおよびテーブルの構成例を示す図である。

メインメモリ 152 のプログラム領域 401 は、オペレーティングシステム 402、実／仮想ボリューム管理 403、バックアップ機能処理 404、装置管理 406、スイッチまたは管理装置との通信処理 408 等から構成され、テーブル領域 153 は、実／仮想ボリューム管理 403 およびバックアップ機能処理 404 と関係するボリューム管理テーブル 405、装置管理 406 と関係する装置管理テーブル 407 等から構成される。

#### 【0010】

図 5 は、管理装置 160 にあるプログラムおよびテーブルの構成例を示す図である。

メインメモリ 501 のプログラム領域 502 は、オペレーティングシステム 504、ストレージ管理ソフト 505 として装置管理 508 や実／仮想ボリューム管理 506、スイッチまたは連携サーバとの通信処理 510 等から構成され、テーブル領域 503 は、実／仮想ボリューム管理と関係するボリューム管理テーブル 507、装置管理 508 と関係する装置管理テーブル 509 等から構成される。

#### 【0011】

図 2 から図 5 で示したボリューム管理テーブル 208、306、405、507 は、図 6 から図 8 に示すユニット管理テーブル 600、論理ユニット管理テーブル 701、仮想ボリューム管理テーブル 702 から構成される。

図 1 で示した計算機システムの管理者が管理装置 160 により、ボリューム管理テーブル 507 を作成または更新し、スイッチ 100 のスイッチ管理部 140 のボリューム管理テーブル 208、連携サーバ 150 のボリューム管理テーブル

405へ配布し、スイッチ管理部140はルーティング制御部120のボリューム管理テーブル306へ配布する。

#### 【0012】

図6は、ユニット管理テーブルの構成例を示す図である。

ストレージ装置180は計算機170やスイッチ100と接続するための複数の接続ポート601、複数のディスク604をアクセスするためのディスクアダプタ603を有し、複数の接続ポート601と複数のディスクアダプタ603がある場合は、内部スイッチ602で各接続ポートと各ディスクアダプタが接続されている。

ユニット管理テーブル600は、複数のエントリ605から構成される。

ディスク毎のエントリ605は、ユニットID、MAC(Media Access Control)アドレス、IPアドレス、ポートID、ディスク番号、ブロック数(説明を簡単にするため、1メガバイトを1ブロックとしている)、容量等の情報を持つ。

#### 【0013】

図7は、論理ユニット管理テーブルと仮想ボリューム管理テーブルの構成例を示す図である。

論理ユニット管理テーブル701は、複数のエントリ701Aから構成される。

論理ユニット毎のエントリ701Aは、論理ユニット(LU)ID、図6で示したユニットID、開始LBA(Logical Block Address)、終了LBA、サイズ、状態等の情報を持つ。

仮想ボリューム管理テーブル702は、複数のエントリ703から構成される。

仮想ボリューム毎のエントリ703は、仮想ボリュームID、構成論理ユニット(LU)の組、サイズ等の情報と、移動可否を示す情報を持ち、エントリ703Aでは移動可否は禁止となっており、エントリ703Bでは移動可否は可能となっている。

一例として示されている仮想ボリューム704は、仮想ボリュームテーブル7

02のエントリ703Aが示すように、論理ユニットIDがVLU00とVLU01の論理ユニットの結合として構成されており、移動可否は禁止である。

#### 【0014】

図8は、図7と同様に論理ユニット管理テーブルと仮想ボリューム管理テーブルの構成例を示す図である。

仮想ボリューム804は、仮想ボリュームテーブル702のエントリ703Bが示すように、論理ユニットIDがVLU10とVLU11の論理ユニットの結合として構成されており、移動可否は可能である。

図7と図8の仮想ボリュームは、同じストレージ装置180内のディスク604で構成されているものが図7、別のストレージ装置180内のディスク604で構成されているものが図8であり、本発明のデータ移行方法は図8のような場合にデータ移行を行う。

図9は、図8の仮想ボリューム804でデータ移行を行った場合の処理の概要を説明するための図である。

#### 【0015】

図10はスイッチ管理部140のデータ移行処理212がデータ移行を行う契機を得るための処理シーケンスの例を示す図である。

図10の処理シーケンスにおいて、仮想ボリューム、例えば仮想ボリューム804を選択すると（ステップ1001）、次に、選択した仮想ボリュームはデータ移動禁止であるか判定し（ステップ1008）、判定の結果がデータ移動禁止であればデータ移行を行わず、データ移動禁止でなければステップ1002に進む。

ステップ1002で構成LUの調査をする。選択した仮想ボリューム804の場合には、仮想ボリューム管理テーブル702のエントリ703Bにより、構成されている論理ユニットはVLU10とVLU11の結合であることが判る。

ステップ1003では、構成LUのIPアドレスが全一致であるかを判定し、全一致であればデータ移行を行わず、全一致でなければステップ1004に進む。仮想ボリューム804の場合には、論理ユニット管理テーブル701のエントリ801A1と801A2によりユニットIDはそれぞれ、RUA0100とRUB

0000であり、ユニット管理テーブル600によりIPアドレスが全一致しないので、ステップ1004に進む。

ステップ1004では構成LUを選択し、ステップ1005に進む。ここでは仮想ボリューム804の場合、RUA0100の論理ユニットを選択したとする。

ステップ1005では、選択した構成LUの記憶装置内に別の構成LUからデータ移行可能な実ユニット内の未使用領域が存在するか判定し、存在すればステップ1007に進み、存在しなければステップ1006に進む。仮想ボリューム804の場合、RUA0100の論理ユニットが選択されており、ステップ1005の判定の結果、別の構成LUであるRUB0000からのデータ移行可能な領域が存在するのでステップ1007に進む。

ステップ1007ではデータ移行処理を行う。仮想ボリューム804の場合、RUB0000内のVLU11をRUA0100へデータ移行する。

ステップ1006では、別の構成LUがあればステップ1004に戻り、無ければデータ移行は行わない。

#### 【0016】

図10に示す処理シーケンスは、データ移行を行わない場合や（ステップ1003のYESの場合）、行うことができない場合もある（ステップ1006のNOの場合）。

また、図10に示す処理シーケンスは、開始の指示を管理装置等からの管理者の指示により行う場合や、ストレージ管理ソフト505のスケジュールにより行う場合もある。

#### 【0017】

図11はデータ移行処理1007の処理シーケンスの例を示す図である。

図12は仮想ボリューム内の論理ユニットの状態を管理するテーブル1200の例を示す図である。図において状態1206はデータ移行に係らないときはアイドルとなる。

図10で示したデータ移行を行う契機を得るための処理シーケンスにおいて、ユニットRUA0100にデータ移行可能な移行先があるということで、論理ユニットVLU11と同じサイズの論理ユニットVLU12をエントリ1209の

ように設定する。その際に状態フィールド 1206 を移行先と設定しておく（ステップ 1101）。

移行元の論理ユニット VLU11 は、エントリ 1208 に示すように状態フィールド 1206 を移行元にする（ステップ 1102）。

ここまでの処理は図 9 の（1）LU 作成である。

次に移行元の論理ユニット VLU11 から移行先の論理ユニット VLU12 にデータを移行し（ステップ 1103）、仮想ボリュームを構成する論理ユニットを移行元の論理ユニット VLU11 から移行先の論理ユニット VLU12 へと変更する（ステップ 1105）。

ここまでの処理は図 9 の（2）データ移行と（3）論理ユニット切替えである。

次に移行先の論理ユニット VLU12 の状態をエントリ 1210 に示すように状態フィールド 1206 をアイドルと設定し、また、仮想ボリューム VVOL02k 移行可否を禁止に設定し（ステップ 1106）、移行元の論理ユニット VLU11 を削除し（ステップ 1107）、移行元論理ユニット VLU11 内のデータを削除する（ステップ 1108）。

ここまでの処理は図 9 の（4）論理ユニット VLU11 削除と（5）論理ユニット VLU11 内データ削除である。

なお、（4）の論理ユニット VLU11 削除と（5）の論理ユニット VLU11 内データ削除とをせずに、一時的に、または指定した期間だけ論理ユニット VLU11 と論理ユニット VLU11 内データをバックアップ用として残しておくようにしてもよい。

#### 【0018】

図 11 のデータ移行処理（ステップ 1103）は、データ移行の最中の Read や Write の I/O も処理する必要があるため、図 13 に示すように、データ移行をブロック単位毎にデータ移行の未完を管理するブロック bitmap テーブル 1301 を有する。

移行元の論理ユニット VLU11 から移行先の論理ユニット VLU12 にブロック単位でデータを移動し（ステップ 1104A）、ブロック bitmap テー

ブル 1301 を更新し (ステップ 1104B)、すべてのデータの移動が完了するまで同様の処理を繰り返す (ステップ 1104C)。

データ移行の最中の I/O 処理 (ステップ 1109) は、図 13 に示すような処理マトリックス 1302 のような処理を行う。

非データ移行領域に対する I/O 処理 1303 は、該当する記憶領域へそのまま適用され、データ移行領域に対する I/O 処理 1304 または 1305 は、ブロック bitmap テーブル 1301 の状態によりデータ移行元またはデータ移行先の論理ユニットの記憶領域へ適用される。

#### 【0019】

以上の図 10 から図 13 で示したスイッチ管理部 140 のデータ移行処理 212 により、ボリューム管理テーブル 208 の構成に変更があった場合は、ルーティング制御部 120 のボリューム管理テーブル 306 や、さらに管理装置 160 のボリューム管理テーブル 507 や、連携サーバ 150 のボリューム管理テーブル 405 にも変更を行う。

#### 【0020】

次に本発明の第二の実施形態を図 14 から図 16 を参照して説明する。

第二の実施形態では、仮想ボリュームの構成に使用されていない比較的小さな容量の記憶領域が少なくなるように、既に仮想ボリュームを構成する一つ以上の記憶領域を別の記憶領域へと、運用中にデータ移行を行う場合について説明する。

図 14 は、論理ユニット管理テーブルと仮想ボリュームテーブルの構成例を示す図であり、図 7 の構成例と同じである。

図 15 はデータ移行を行う契機を得るための処理シーケンスの例を示す図である。

図 16 はユニットの構成容易度を管理するユニット使用状況管理テーブル 213 の例を示す図である。

構成容易度は、一例として空きブロック数が 7,500 未満の場合は「低」、12,500 未満の場合は「中」、12,500 以上の場合は「高」としている。



## 【0021】

図15の処理シーケンスにおいて、まず、ユニットの選択をする（ステップ1501）。例として、ユニットRUA0001（図6、図7のエントリ604A2）を選択したとする。

次に、記憶領域の構成容易度を調査し（ステップ1502）、ステップ1503に進む。

ステップ1503では、選択したユニットに構成容易度の低い領域があるか判定し、無ければデータ移行は行わず、有ればステップ1504に進む。例の場合、ユニットRUA0001の各記憶領域の構成容易度を調査し、図16のユニットIDがRUA001のエントリには構成容易度のフィールド1605に「中」と「低」の記憶領域があることが判るので、図10の処理シーケンスで示した処理ステップ1004以降と同様の処理を、図15の処理ステップ1504以降の処理で行う。

ただし、第二の実施形態では、移行元の論理ユニットのデータを最終的な移行先の論理ユニットへ直接移行できない場合に、少なくとも一つの間接的な論理ユニットに一旦データ移行し、その後に最終的な移行先の論理ユニットにデータ移行する場合を示している。

## 【0022】

図14において、（1）ユニットRUA000Xに間接的な論理ユニットVLU02の作成を行い、（2）ユニットRUA0000にある移行元の論理ユニットVLU01からユニットRUA000XにあるVLU02にデータをコピーした後、ユニットRUA000XにあるVLU02のコピーしたデータを最終的な移行先であるユニットRUA0000にコピーし、このコピーしたデータに対して新たな論理ユニットVLU01を設定し、（3）元の論理ユニットVLU01を新たな論理ユニットVLU01に切替え、（4）間接的な論理ユニットVLU02の削除と（5）論理ユニットVLU02内のデータ削除を行う。

図16において、ユニットRUA000Xは図14の（1）でエントリ1607のように変更され、ユニットRUA0001とRUA000Xは図14の（4）でエントリ1608のように変更され、最終的にユニットRUA0001は構

成容易度が高くなっている。

### 【0023】

次に本発明の第三の実施形態を図17と18を参照して説明する。

第三の実施形態では、スイッチの主要構成要素を冗長構成にした場合の説明と、管理装置からのテーブルの配布方法と、スイッチ管理部のテーブルの同期方法について説明する。

図17はスイッチ100の主要構成要素を冗長構成にした場合の計算機システムの構成例である。

図17では、ルーティング制御部120、クロスバースイッチ130、スイッチ管理部140が冗長構成となっている。

第一の実施形態から第二の実施形態で示したような、ユニット管理テーブル600、論理ユニット管理テーブル701、仮想ボリューム管理テーブル702は、図18の情報の流れ1802のように管理装置160からスイッチ管理部140Aと140Bに配布され、メインメモリ142Aと142Bに格納される。

冗長構成のスイッチ管理部140は、現用系と予備系のように動作してもよいし、どちらも現用系で処理分担をしてもよい。

冗長構成のスイッチ管理部140の片系が障害となった場合は、現用系／予備系の場合は予備系に切替わり、どちらも現用系の場合は縮退運用となる。

冗長構成のスイッチ管理部140の持つテーブル143は、情報の流れ1801のように同期をとっておくことにより、片系が障害となった場合に比較的高速な切替えが可能となる。

### 【0024】

次に本発明の第四の実施形態を図19を参照して説明する。

第四の実施形態ではスイッチのスイッチ管理部がハードディスクを持つ構成の場合にデータ移行を行う場合について説明する。

図19は、スイッチ100のスイッチ管理部140がハードディスク1901を有する構成例を示した図である。

第一の実施形態から第二の実施形態で示したように、ハードディスク1901もユニット管理テーブル600（図6参照）の一つとして扱うことにより、論理

ユニットの作成や仮想ボリュームへの構成を可能とする。

ユニット管理テーブル 1 9 0 3 は、一例としてエントリ 1 9 0 4 のような構成をとる。

ユニット R U S 0 0 0 0 にデータ移行の際の間接的な論理ユニットを作成することにより、ストレージ装置 1 8 0 (図 1 参照) の記憶領域が比較的少なくなってきた場合に使用することが可能であり、データ移行時のデータがスイッチを通過する回数を低減することが可能となる。

#### 【 0 0 2 . 5 】

##### 【発明の効果】

本発明によれば、比較的少ない記憶装置からの記憶領域により仮想ボリュームを構成し、中継装置の中継処理負荷を低減させることができる。

また、比較的小さな容量の未使用の記憶領域の低減と、結果として比較的大きな容量の連続した空きの記憶領域の増加を図ることができる。

##### 【図面の簡単な説明】

##### 【図 1】

計算機システムの構成例を示す図である。

##### 【図 2】

スイッチのスイッチ管理部が持つプログラムおよびテーブルの構成例を示す図である。

##### 【図 3】

スイッチのルーティング制御部が持つプログラムおよびテーブルの構成例を示す図である。

##### 【図 4】

スイッチの連携サーバが持つプログラムおよびテーブルの構成例を示す図である。

##### 【図 5】

計算機システムの管理装置が持つプログラムおよびテーブルの構成例を示す図である。

##### 【図 6】

ユニットの構成例とユニット管理テーブルの構成例を示す図である。

【図 7】

論理ユニット管理テーブルと仮想ボリューム管理テーブルの構成例を示す図である。

【図 8】

論理ユニット管理テーブルと仮想ボリューム管理テーブルの構成例を示す図である。

【図 9】

データ移行の処理概要を示す図である。

【図 10】

データ移行の契機を得るための処理シーケンスの例を示す図である。

【図 11】

データ移行の処理シーケンスの例を示す図である。

【図 12】

論理ユニットの状態を管理するテーブルの例を示す図である。

【図 13】

データ移行中の I/O 処理を行うためのブロック bitmap テーブルと I/O 処理マトリックスの例を示す図である。

【図 14】

データ移行の処理概要を示す図である。

【図 15】

ディスクの構成容易度をデータ移行の契機にする場合の処理シーケンスの例を示す図である。

【図 16】

ユニット使用状況管理テーブルの構成例を示した図である。

【図 17】

スイッチの主要構成要素を冗長構成とした場合の構成例を示した図である。

【図 18】

冗長構成のスイッチ管理部のテーブル情報の配布および同期の方法を示す図で

ある。

【図 19】

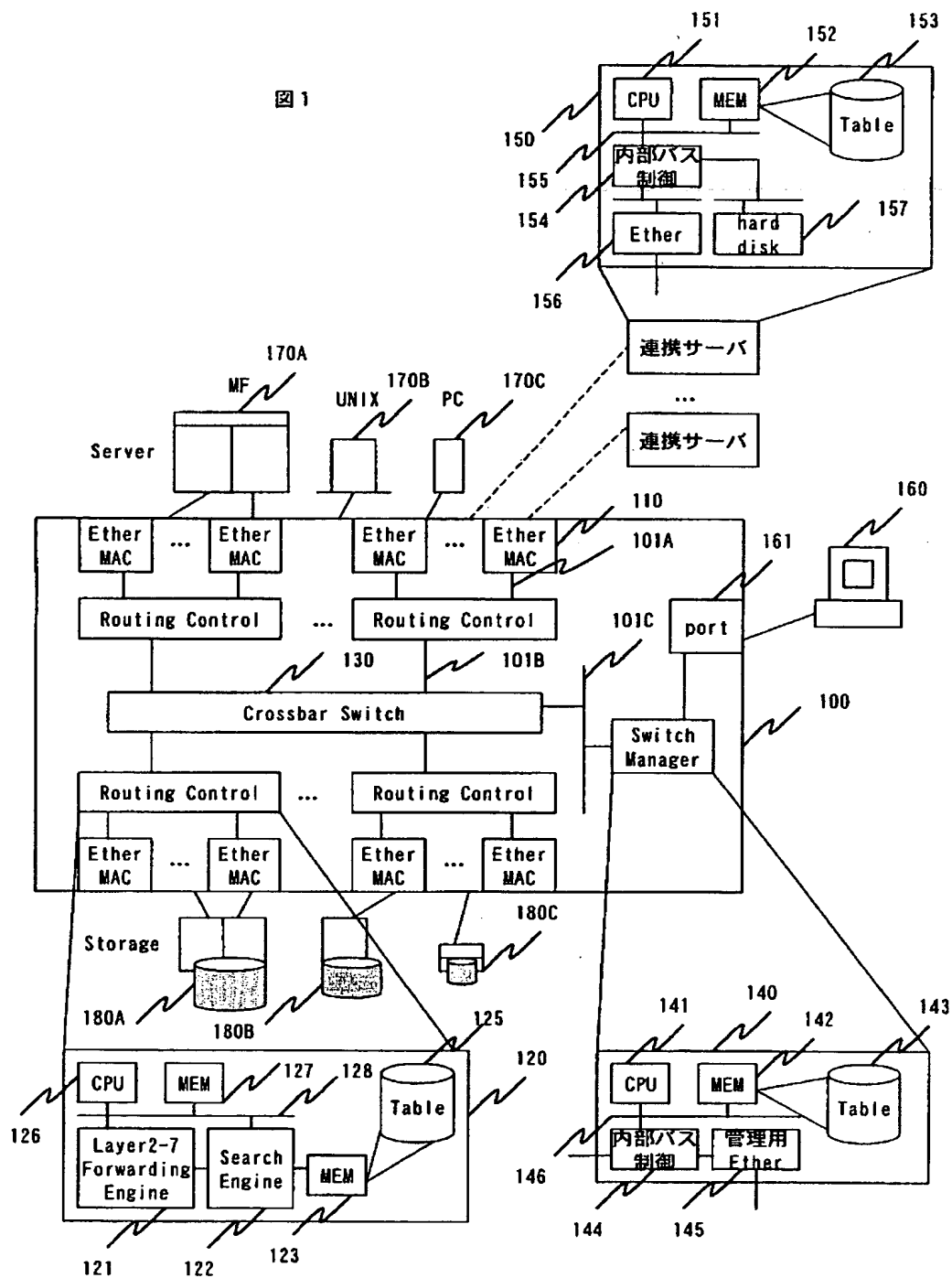
スイッチ管理部がハードディスクを持つ構成例を示す図である。

【符号の説明】

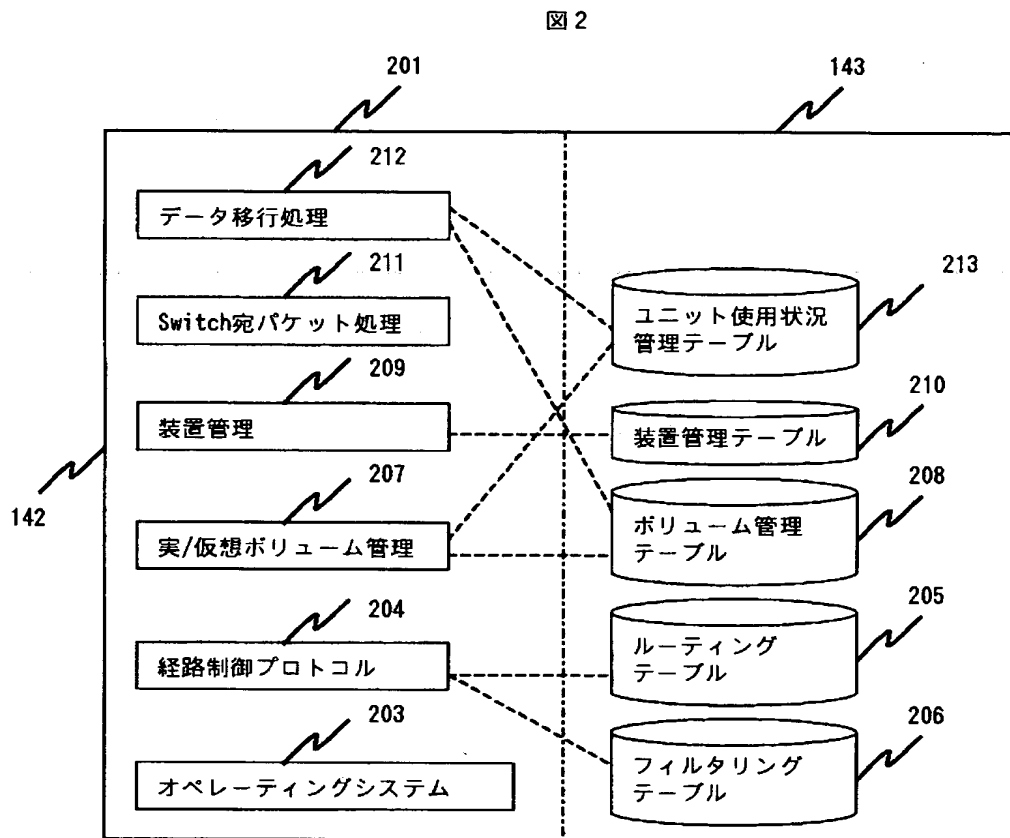
- 100 スイッチ
- 110 接続インタフェース
- 120 ルーティング制御部
- 130 クロスバースイッチ
- 140 スイッチ管理部
- 150 連携サーバ
- 160 管理装置
- 161 ポート
- 170 サーバ装置
- 180 ストレージ装置

【書類名】 図面

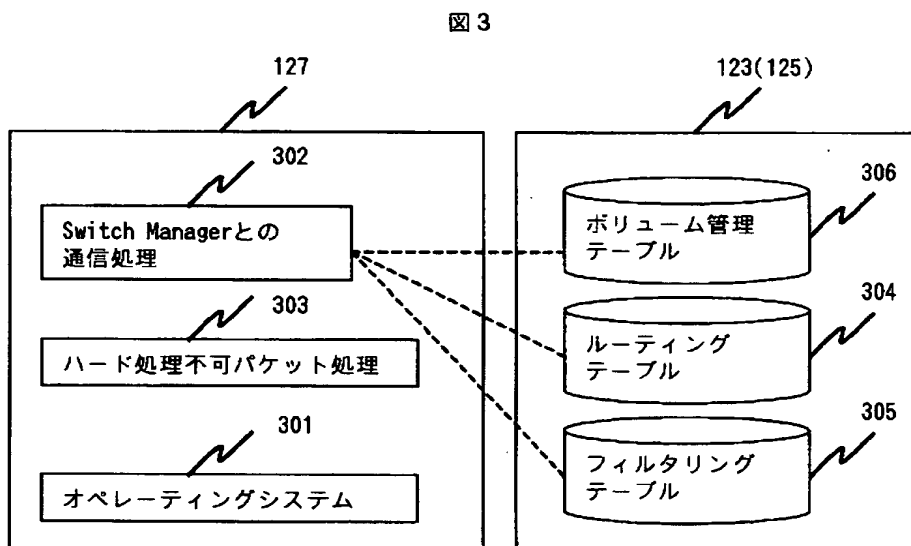
【図 1】



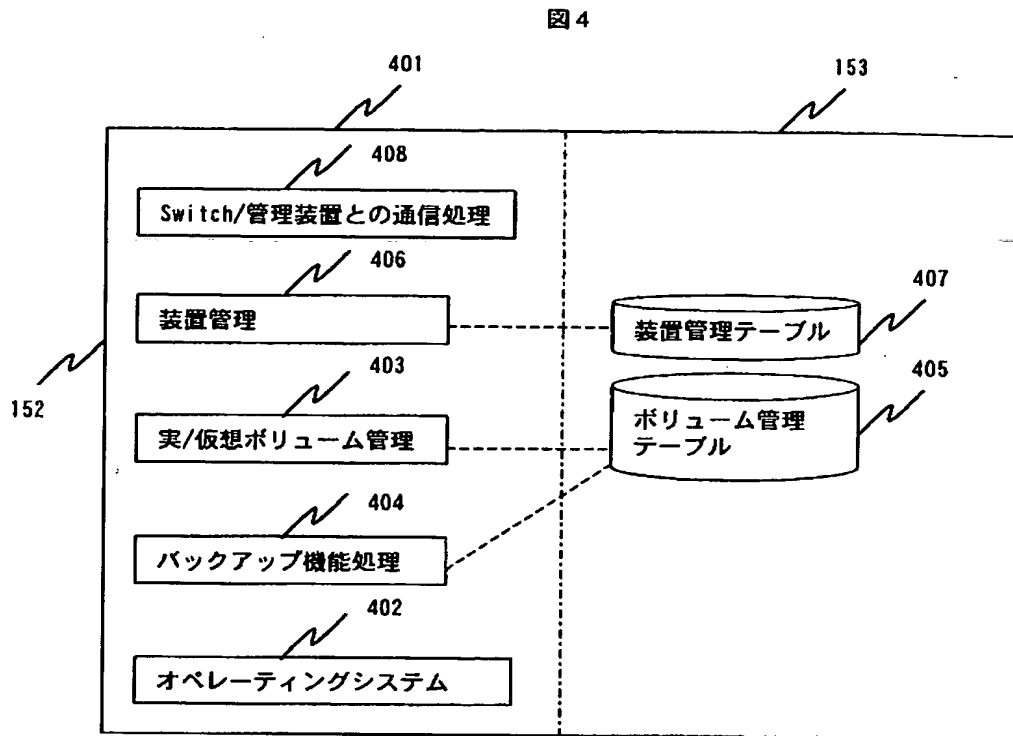
【図 2】



【図 3】

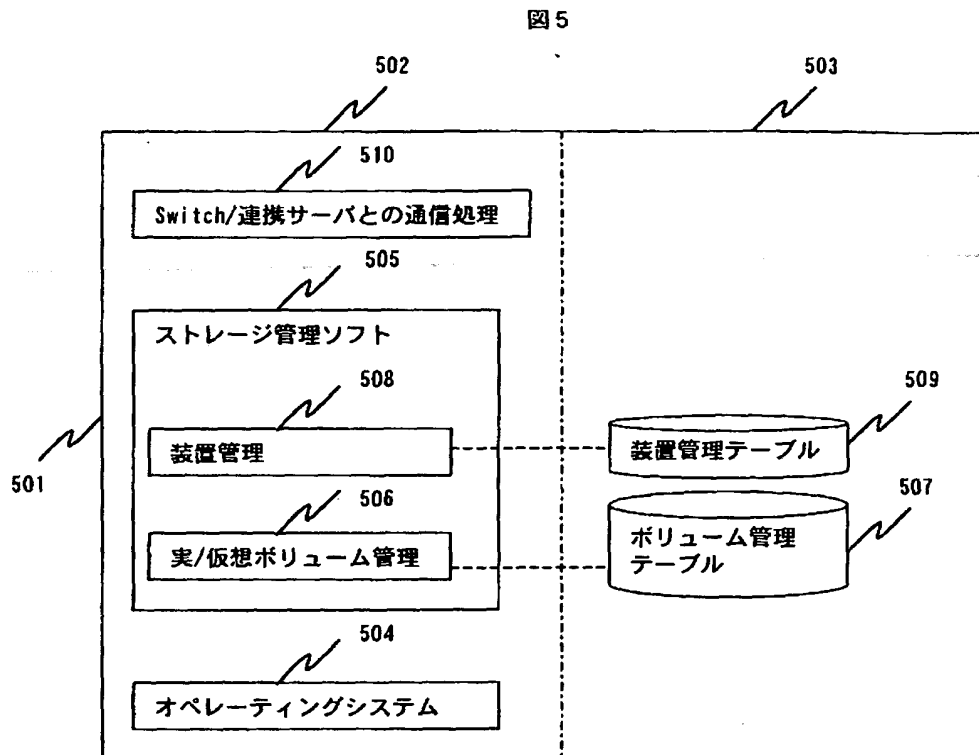


【図 4】



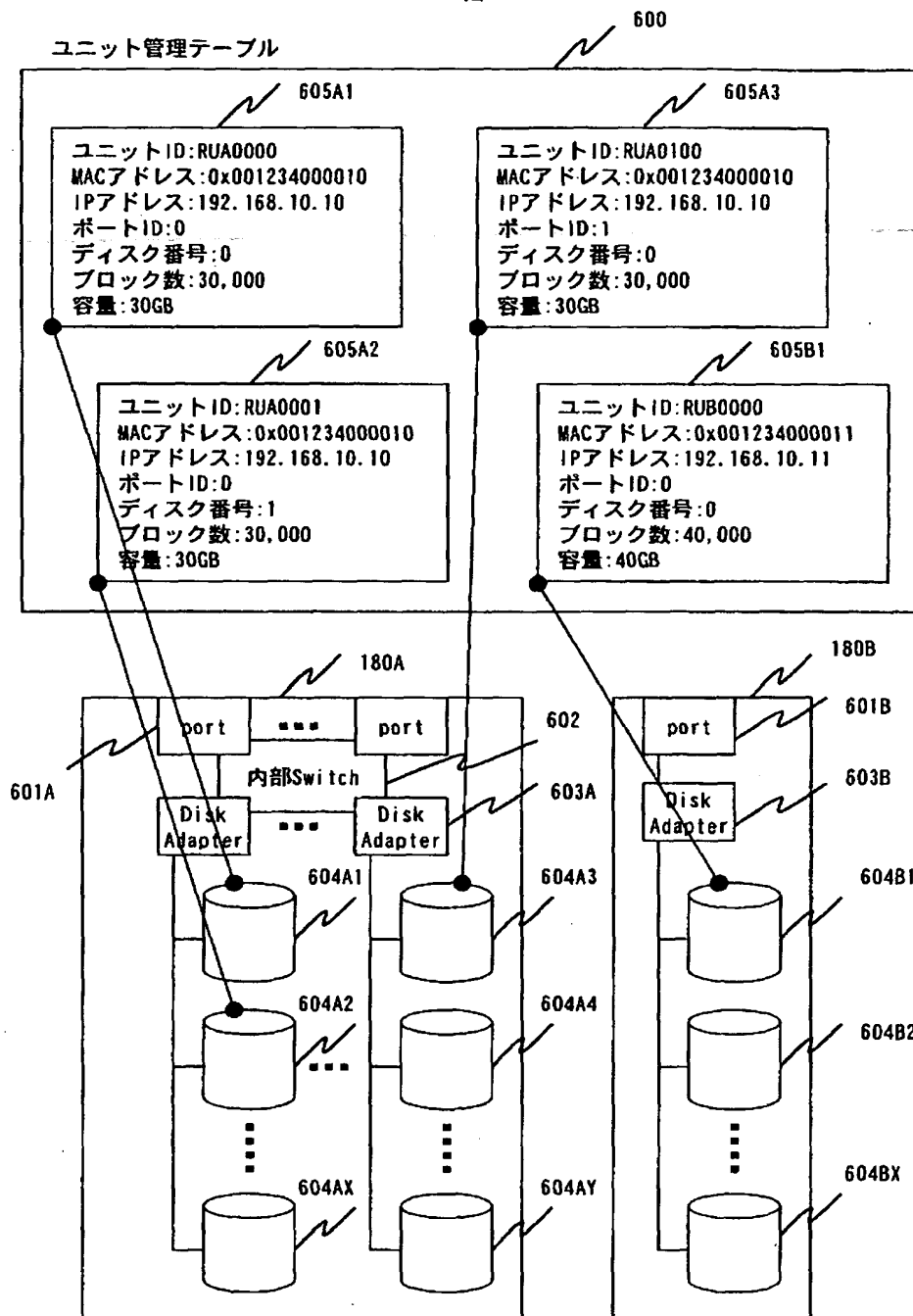


【図 5】

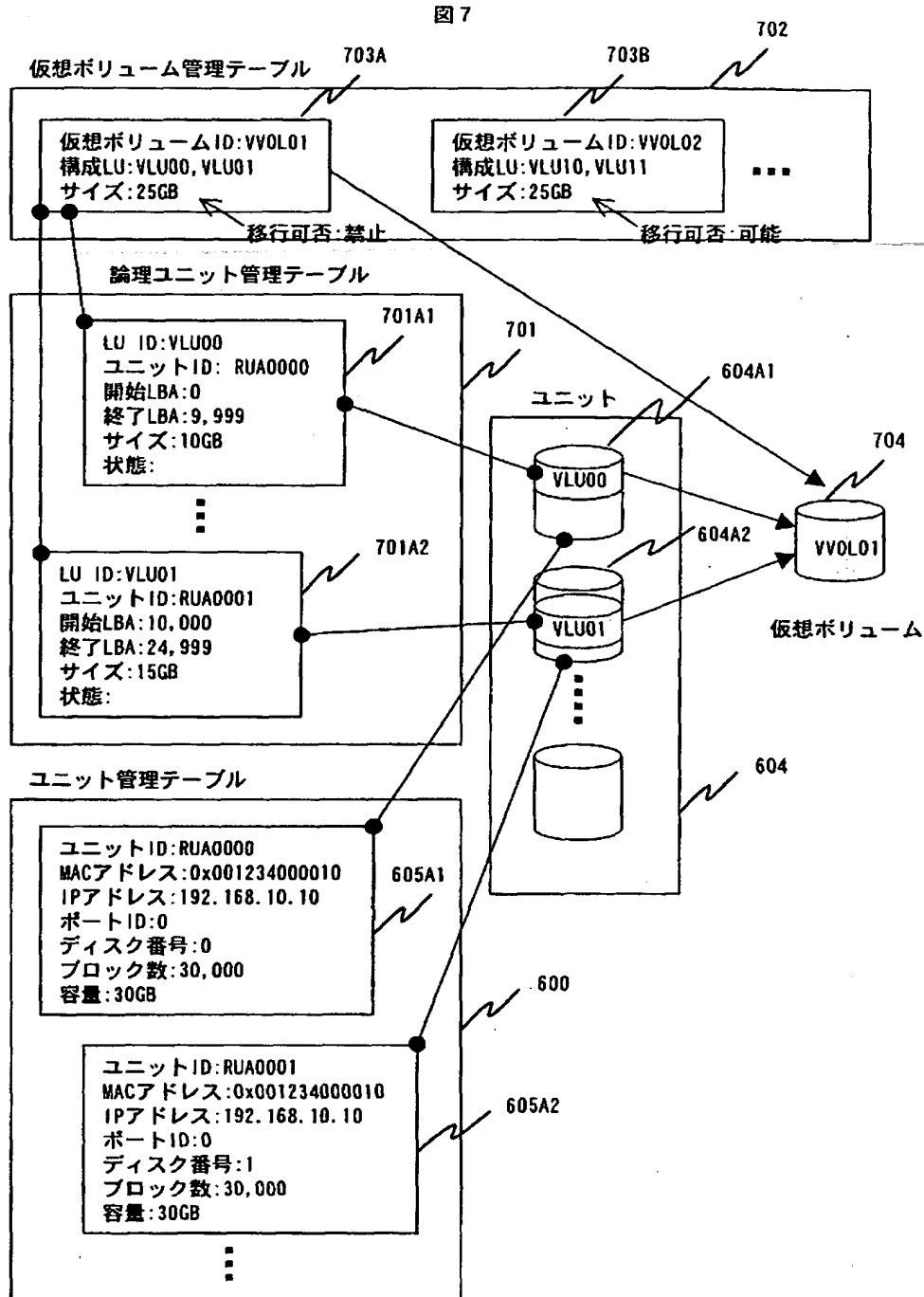


【図6】

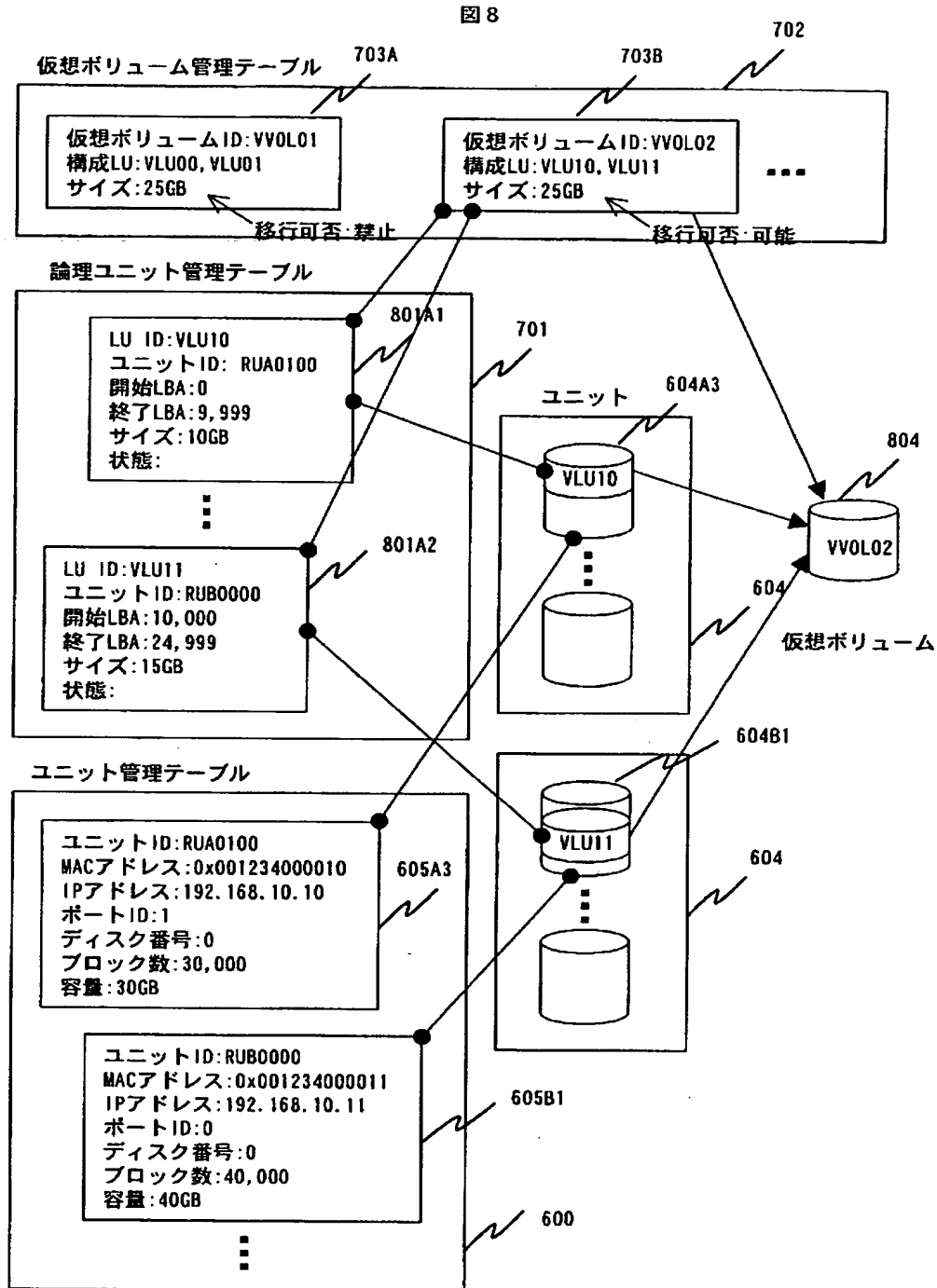
図6



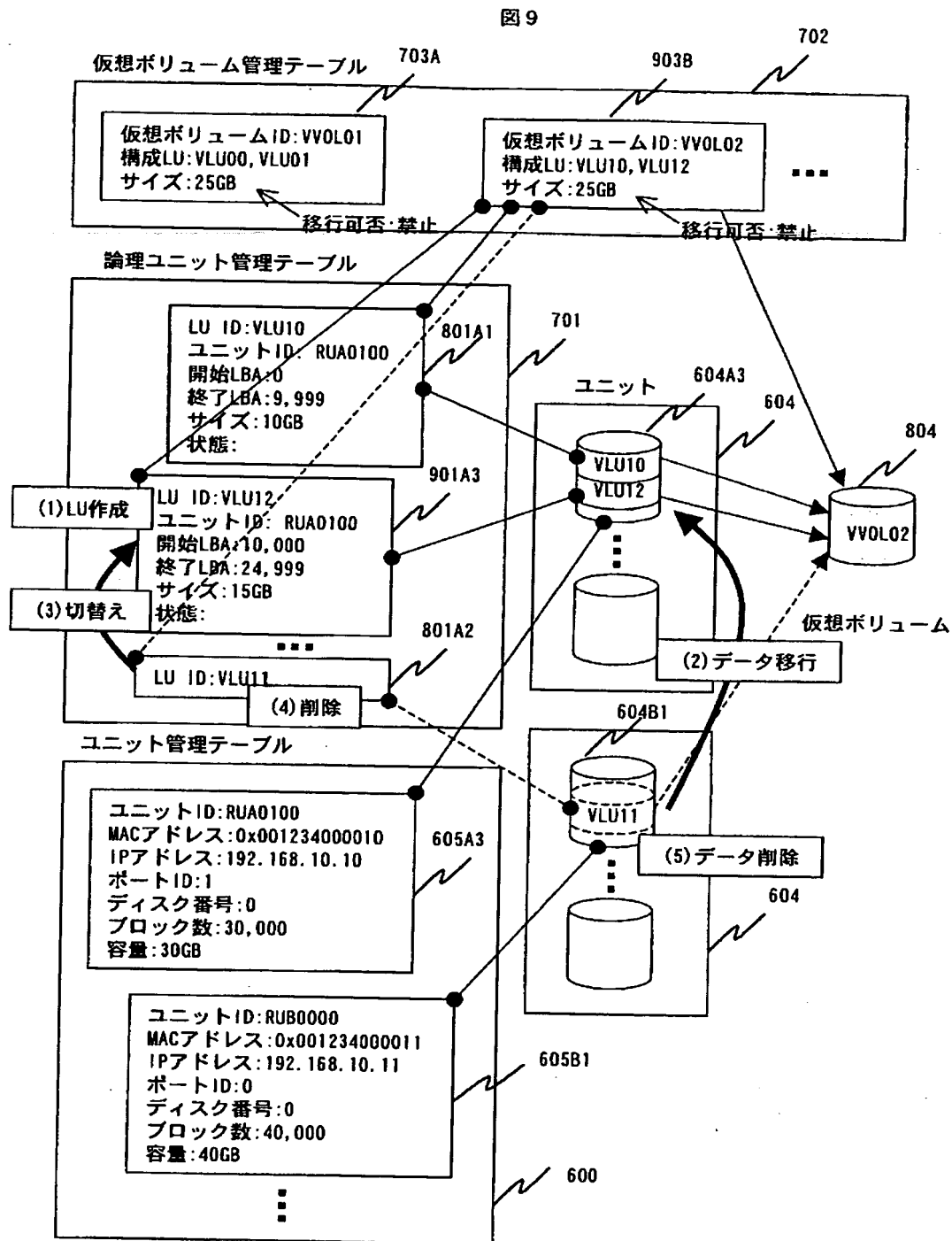
【図7】



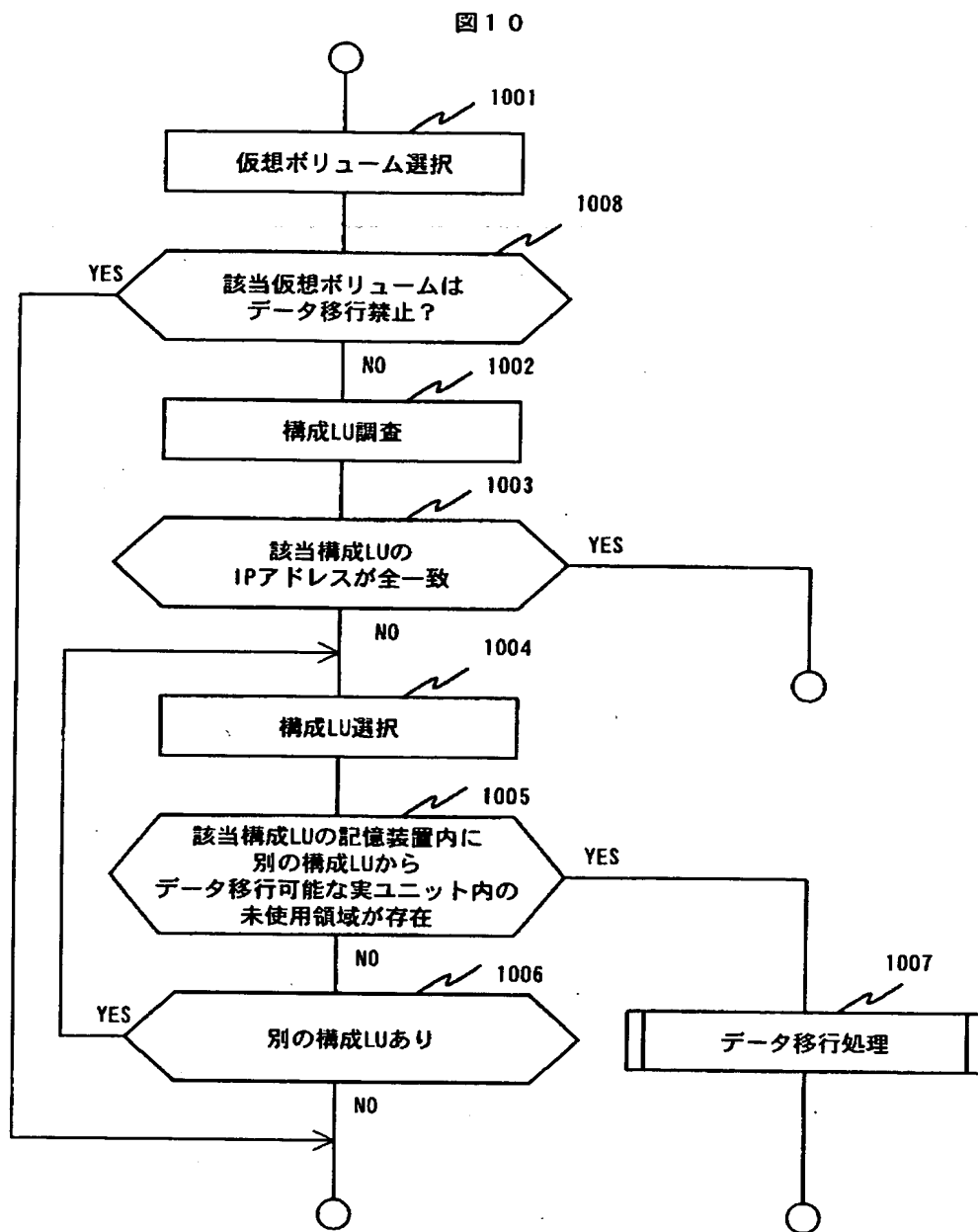
【図 8】



【図9】

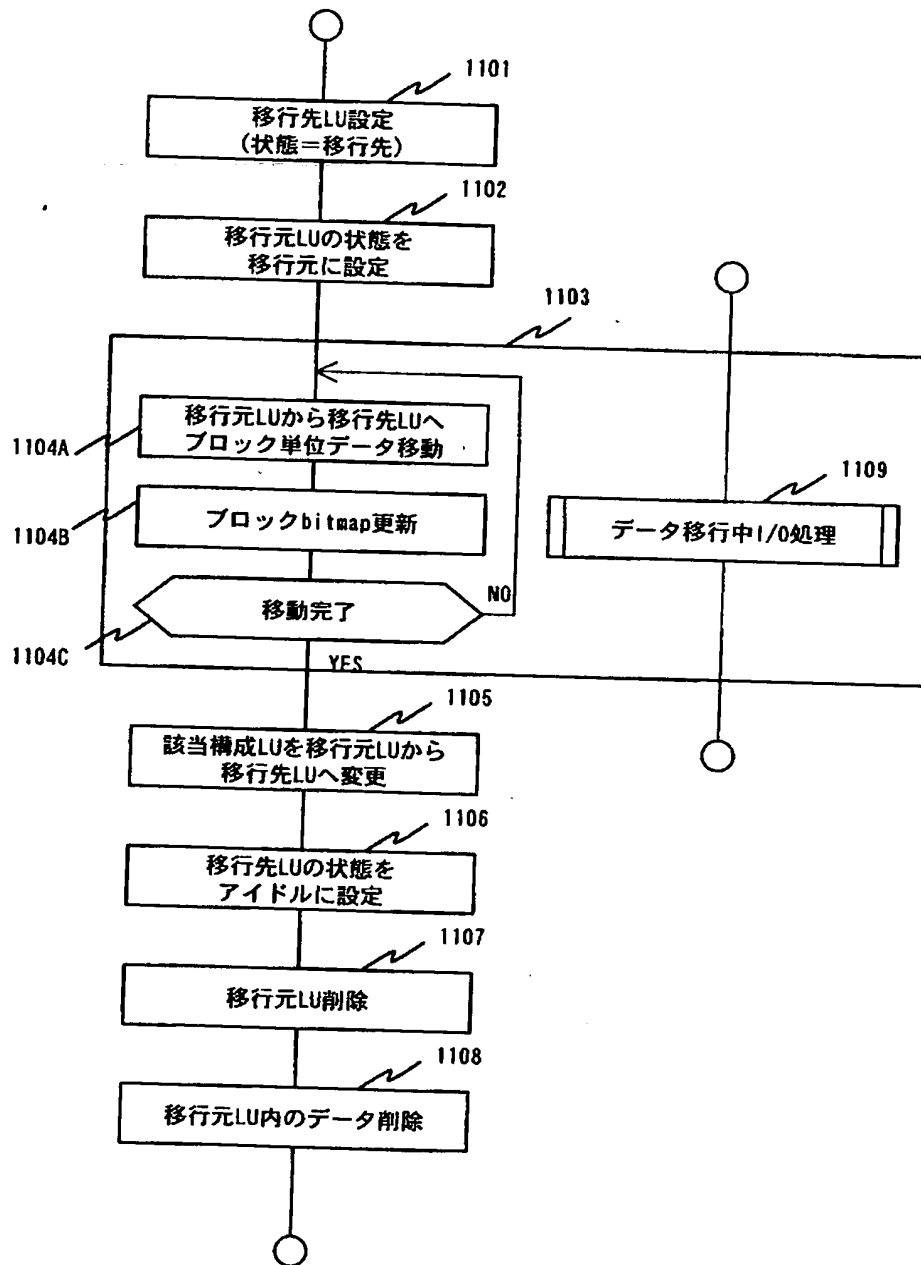


【図 10】



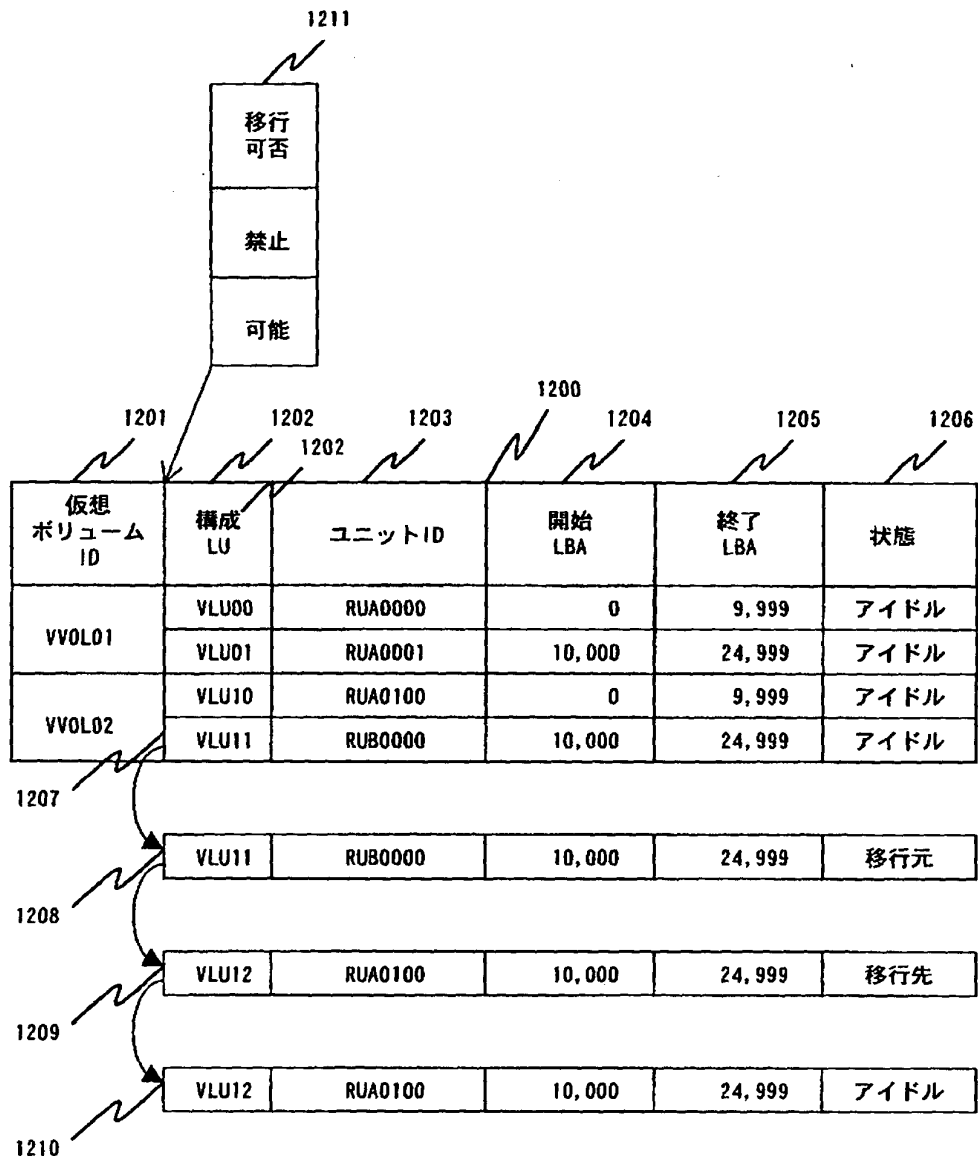
【図11】

図11



【図 12】

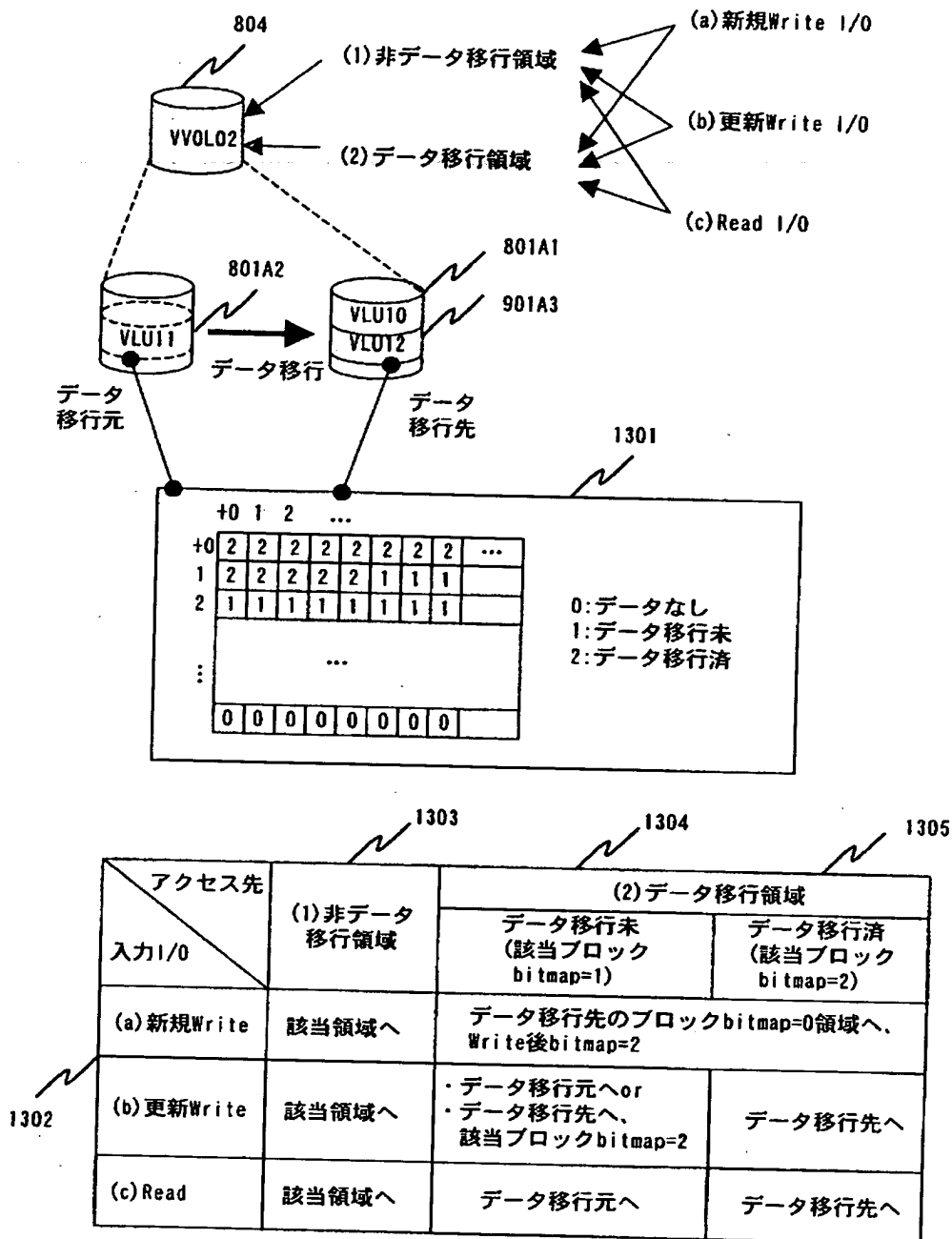
図 12



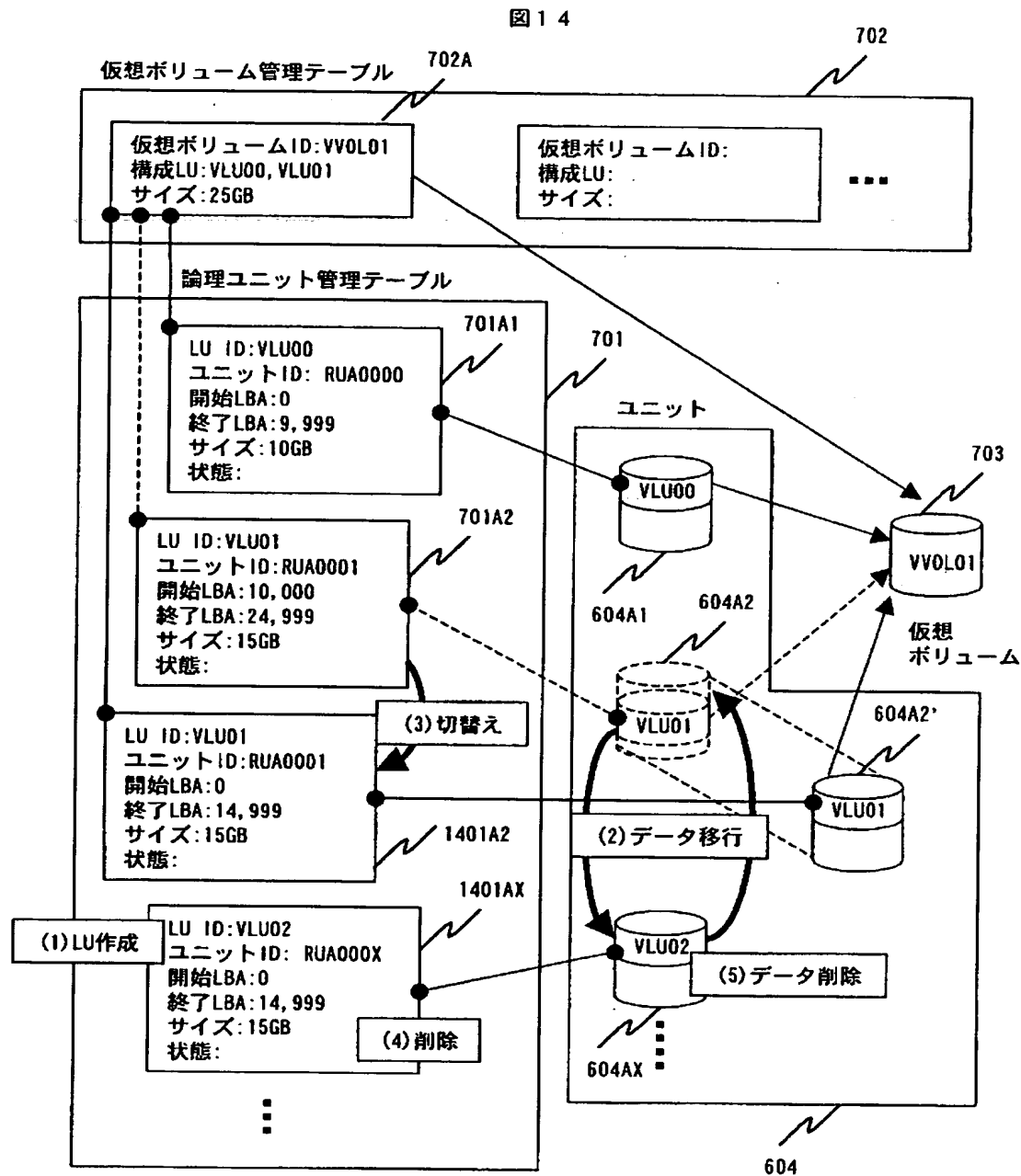


【図13】

図13

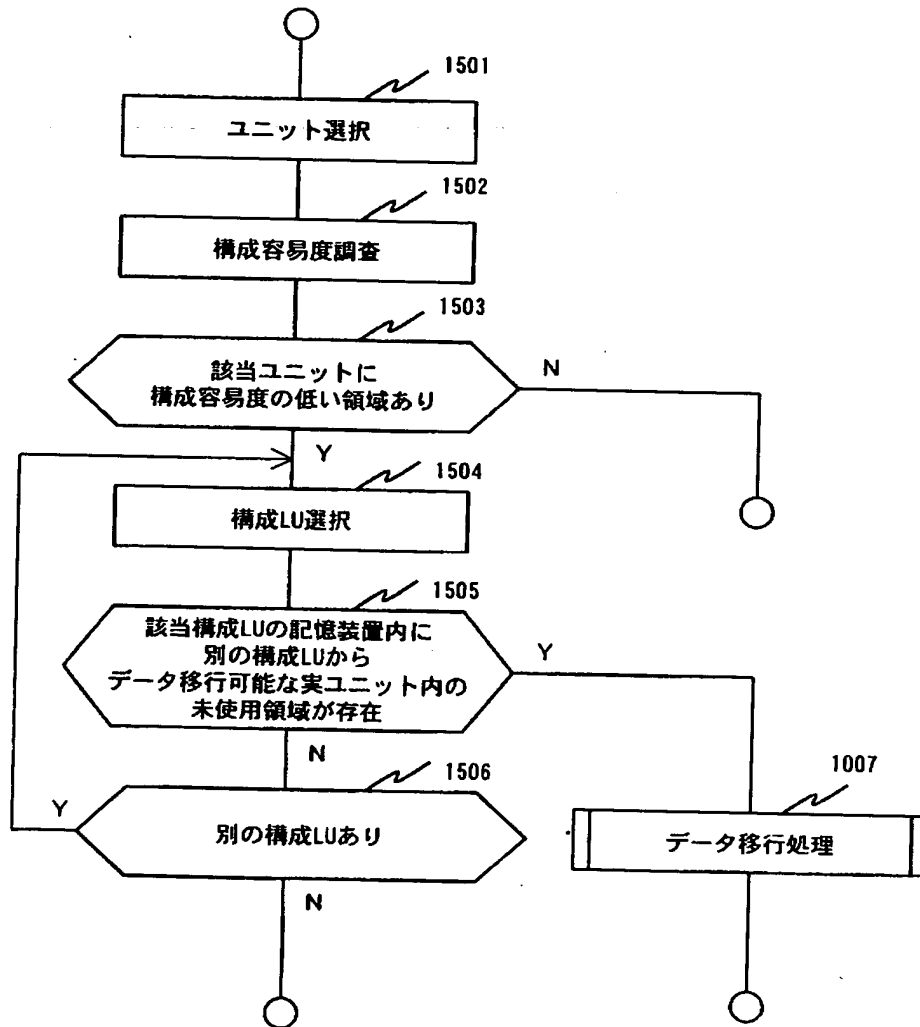


【図14】



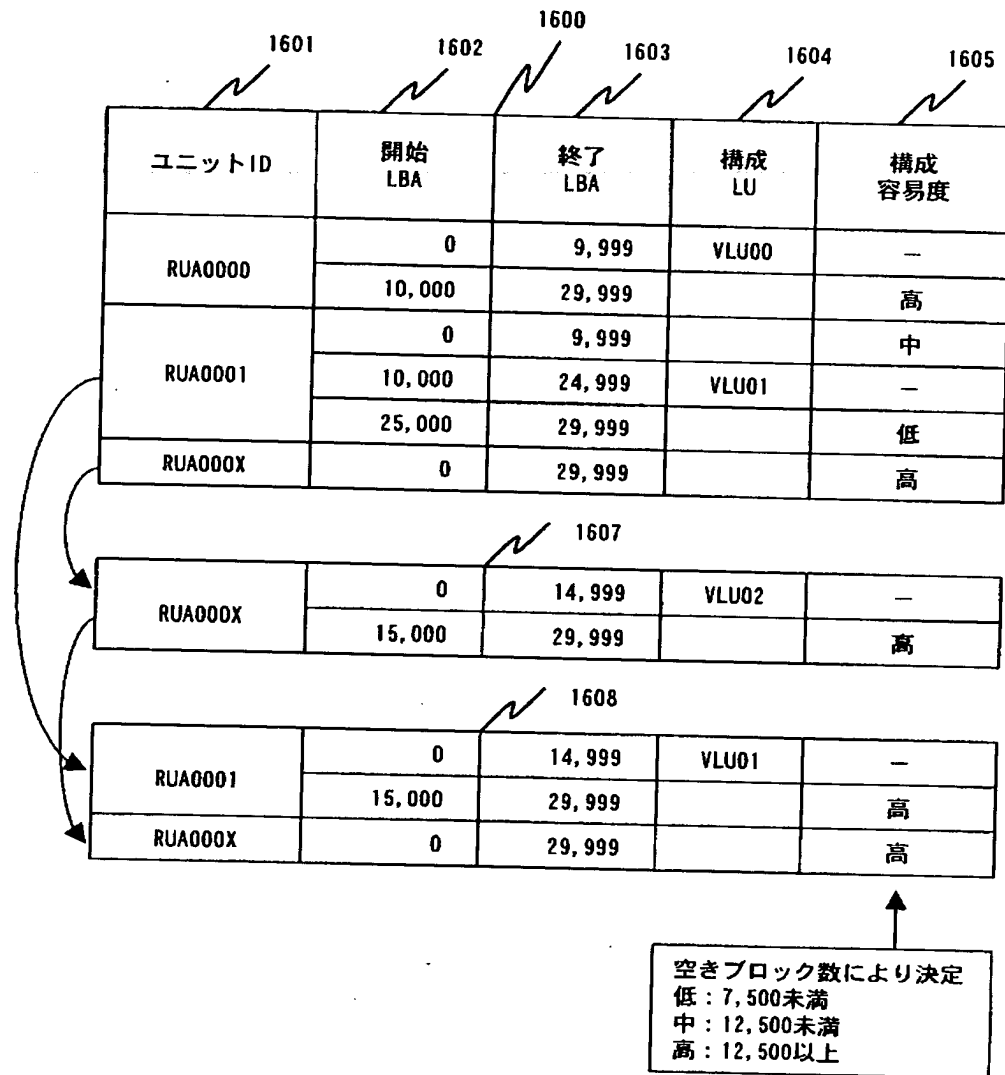
【図 15】

図 15



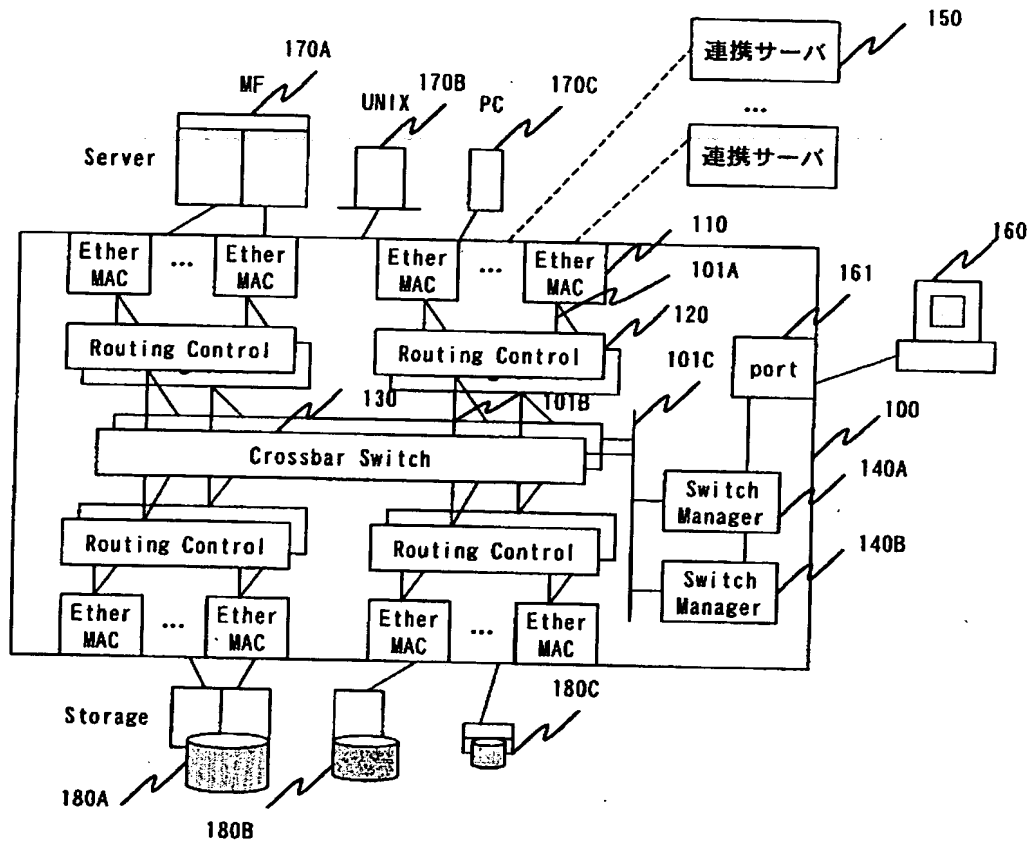
【図 16】

図 16



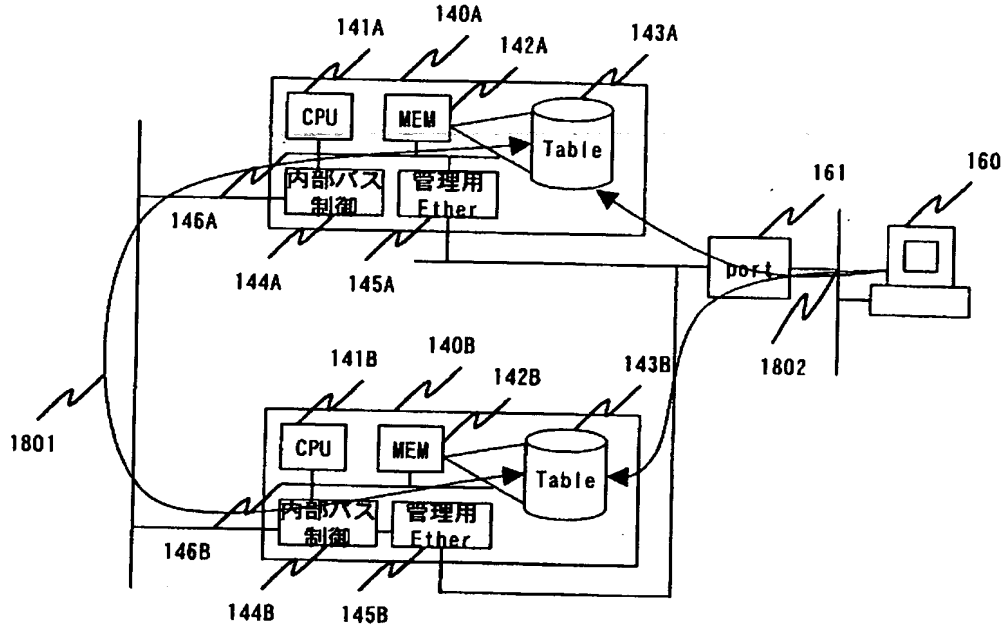
【図 17】

図 17



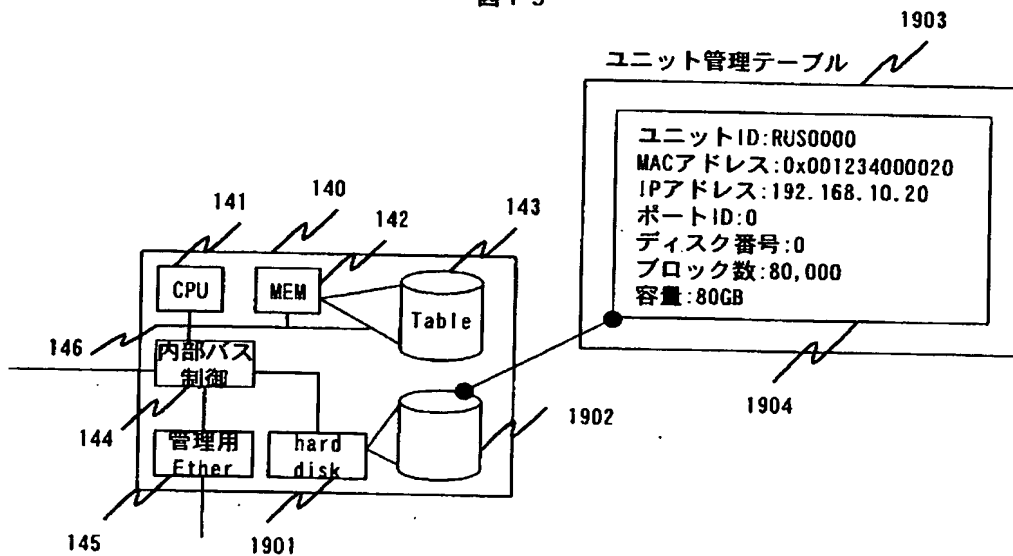
【図 18】

図 18



【図 19】

図 19



【書類名】 要約書

【要約】

【課題】 比較的少ない数の記憶装置からの記憶領域により仮想ボリュームを構成するように、中継装置が運用中にデータ移行を行うことにより中継処理負荷を低減させる方法を提供する。

【解決手段】 ある一つの仮想ボリュームを構成している記憶装置（180）の宛先が増加してくると、中継装置（100）がこれを契機にできるだけ少ない記憶装置の宛先になるように仮想ボリュームの構成を変更できるように、複数の記憶装置間でデータ移行を行う。また、仮想ボリュームを構成していない未使用の記憶領域の中で比較的小さな容量の記憶領域が増加してくると、中継装置がこれを契機にできるだけこの小さな容量の記憶領域の数を低減するように、複数の記憶装置間でデータ移行を行う。

【選択図】 図 1

特願 2003-038097

出願人履歴情報

識別番号

[000005108]

1. 変更年月日  
[変更理由]  
住所  
氏名

1990年 8月31日  
新規登録  
東京都千代田区神田駿河台4丁目6番地  
株式会社日立製作所